NUCLEAR INSTRUMENTS AND METHODS 50 (1967) 309-320; © NORTH-HOLLAND PUBLISHING CO.

A METHOD FOR AUTOMATIC IDENTIFICATION OF PEAKS IN THE PRESENCE OF BACKGROUND AND ITS APPLICATION TO SPECTRUM ANALYSIS*

M.A. MARISCOTTI**

Brookhaven National Laboratory, Upton, New York, U.S.A.

Received 15 August 1966

A computer method is described that permits automatic analysis of high resolution spectra, provided that the data are stored in a way suitable for computer input (i.e., on magnetic or paper tape). The program reads the entire spectrum, locates the peaks and determines their amplitudes, centroids and widths by least squares fitting. Doublets are handled automatically.

1. Introduction

Before the solid state detectors were introduced in nuclear spectroscopy, a large number of methods and computer codes had been developed for the analysis of nuclear spectra obtained with NaI(Tl) scintillation detectors[1]). The correct evaluation of the different components contributing to the composite spectrum represented in these cases the principal problem to be solved. The assumption of a standard spectrum corresponding to a given gamma-ray line was necessary for applying either the "stripping" technique, consisting of successive subtraction of the individual components of the spectrum, or a least squares fitting procedure to deduce the contribution of each component to the total spectrum.

At the present time, the high resolution of the Ge(Li) detectors has somewhat simplified the analysis of a gamma-ray spectrum. Since the width of a peak represents a very small part of the whole spectrum, a smooth background may usually be assumed under each line. Gamma ray energies and intensities can then be simply obtained by means of a least squares fitting of each line and its adjacent channels which define the local background.

On the other hand, the high resolution has led to the application of large multichannel analyzers and the spectra obtained with equipment of this type usually contain many peaks. For this reason the analysis of a whole spectrum may still demand a considerable amount of work although the study of each peak is simpler. It is necessary to locate the lines throughout the spectrum, then to determine for each one the group of channels that should undergo fitting and finally – the most time consuming part – to punch out the data contained in these channels to feed the least squares program.

The purpose of the method described here is to avoid the delay and effort associated with these separate steps. It makes it possible to use the computer to search for peaks in the presence of a background and statistical fluctuations and to select for each peak the suitable channels before entering a standard least squares fitting subroutine. The usefulness of this method becomes apparent if one considers that the raw data are normally stored on magnetic or paper tape and either one is already a suitable input for the computer. Thus, the program goes through the following steps:

1. It reads the entire spectrum from the tape generated by the analyzer.
2. It determines the existing peaks according to the method described below.
3. It finds the Gaussian function which fits each line best using the subroutine of[2]) (a double Gaussian function is employed by the program when it discovers a doublet).
4. It lists the resulting centroids, amplitudes and widths, along with their errors, and plots for each peak the calculated Gaussian function superimposed on the experimental data.

The description of the method and its application is divided in four sections. In section 2 a function is defined that is used to find the peaks. The computer calculates the second difference of the counting rate as a function of the channel number and carries out a "smoothing" of this function (averaging neighboring points) to reduce the effect of the statistical fluctuation. If the two parameters that govern this smoothing procedure are properly chosen, the resulting function tends to vanish everywhere except in the vicinity of a peak.

In section 3 the most suitable choice of these parameters z and w is discussed. Using such a function, as explained in section 4, the computer recognizes the

* Work performed under the auspices of the U.S. Atomic Energy Commission.

** Permanent address: Departamento de Fisica, Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina.

presence of a peak wherever a group of points lying at least one standard deviation above zero precedes another group of negative points and this sequence can be correlated with the second derivative of a Gaussian function.

The last section describes the application of a non-linear least squares fitting subroutine[2]) to fit the identified lines and gives examples of applications.

2. Principle of the method

We will assume that peaks can be described by Gaussian functions and that the background may be approximated by a linear function within short intervals. For a high resolution spectrum the latter assumption is usually quite well fulfilled in an interval which has several times the peak width. Thus, in such an interval, the number of counts as a function of the channel number x is

$$N(x) = G(x)+B+Cx, \tag{1}$$

where $G(x)$ represents a Gaussian function if a peak is present and is zero otherwise. B and C are constants describing the background.

If one assumes that $N(x)$ is a continuous function, its second derivative $N''(x)$ becomes independent of the background and it vanishes for any interval in which there is no peak. Therefore a peak would be located wherever $N''(x) \neq 0$ and on this fact the method is based.

For a real case however, the data N_i represent a discrete function of the channel number i. In analogy to eq. (1) and assuming a peak in the interval considered, we will write

$$N_i = A \exp\left[-(i-i_0)^2/(2\sigma^2)\right]+B+Cx, \tag{2}$$

where A represents the intensity of the peak centered at channel i_0 and B and C are the same as in eq. (1). The parameter σ is related to the peak fwhm Γ by

$$\Gamma = 2.355\,\sigma.$$

Because of the discrete nature of the data, the second derivative is also replaced by the second difference

$$S_i = N_{i+1}-2N_i+N_{i-1}, \tag{3}$$

which, like $N''(x)$, should be only different from zero in the vicinity of a peak. However, since the data N_i are defined within a statistical error, the S_i values will fluctuate around the expected value according to their standard deviation and, if the expected value S_i at $i = i_0$ is comparable to its standard deviation, no peak searching can be attempted. This effect obviously depends on the peak height and width as well as on the background intensity.

A simple calculation shows that S_i is comparable to its standard deviation

$$F_i = (N_{i+1}+4\,N_i+N_{i-1})^{\frac{1}{2}}, \tag{4}$$

even at $i = i_0$, unless A is very large. Let $A_{\min}$ be the peak intensity for which $S_{i_0} = F_{i_0}$. In the specially advantageous case of $B = C = 0$ in eq. (2), one obtains $S_{i_0} \gg F_{i_0}$ only if $A \gg A_{\min} = 6.25\,\sigma^4$. In this case for instance, if $\sigma = 4$, $A_{\min} = 1600$.

Because much weaker peaks should be also resolved, it is necessary to modify the function S_i in order to reduce its standard deviation. For this purpose one can average neighbouring points in order to get a "smoothed" second difference function as expressed by

$$S_i(w) = \sum_{j=i-m}^{i+m} S_j, \tag{5}$$

where we have omitted the normalization constant. The parameter $w = 2m+1$ (m is integer) is to be chosen. According to this definition w is always an odd number.

Now, one can again compare eq. (5) (choosing a suitable w) with its corresponding standard deviation and find a new value of $A_{\min}$. If the latter is smaller, one can hope to reduce its value further by averaging, in turn, the new function $S_i(w)$, and so on, until the smallest $A_{\min}$ is obtained. To find, then, both the number of iterations and the value w for which this minimum occurs we shall consider the following generalized second difference

$$S_i(z, w) = \underbrace{\sum_{j=i-m}^{i+m} \cdots \sum_{h=l-m}^{l+m}}_{z} S_h. \tag{6}$$

This function depends now on two parameters, the number z of averaging steps carried out and the "window" w already defined in eq. (5). The expressions S_i in eq. (3) and $S_i(w)$ in eq. (5) are both particular cases of $S_i(z, w)$, for $z = 0$ and $z = 1$, respectively. This smoothed second difference function may approach sufficiently the ideal function $N''(x)$ if z and w are suitably chosen.

The determination of the optimum pair (z_0, w_0) is the object of the next section. It will be shown that the most appropriate pair is

$$z_0 = 5, \qquad w_0 = 0.6\,\Gamma. \tag{7}$$

With this choice, the quantity $A_{\min}$ is reduced to 0.4 for the case considered above, i.e., $B = C = 0$, $\sigma = 4$ and N_i as in eq. (2). If B assumes a more realistic value, for instance $B = 5000$, we obtain $A_{\min} \cong 50$.

Fig. 1 shows the application of both $S_i(z = 0)$, eq. (3) and S_i ($z = 5$, $w = 5$) to the experimental

spectrum shown uppermost[3]), which has been chosen on account of its large background (and, hence, large statistical fluctuation), the presence of doublets and of both intense and weak peaks. The corresponding standard deviations are also plotted so that the influence of statistics upon both functions can be seen. The thirteen peaks numbered at the top are those identified by the computer using $S_i(z_0, w_0)$, its standard deviation, and the procedure explained in section 4.

The expression (6) is not appropriate for a straightforward computation of the standard deviation. For this reason we will write $S_i(z, w)$ in terms of the data N_j introducing the coefficients C_{ij} such that

$$S_i(z, w) = \sum_{\text{all } j} C_{ij}(z, w)\, N_j. \tag{8}$$

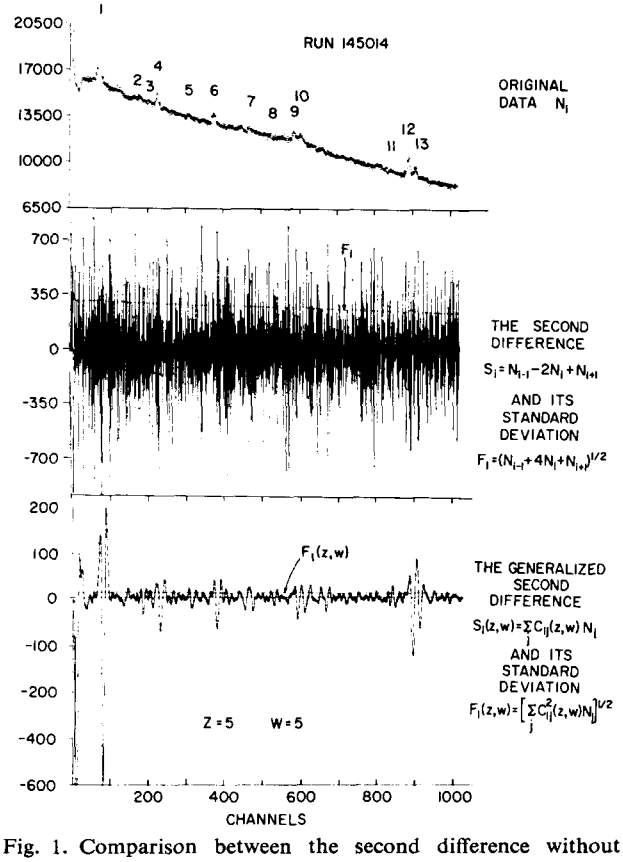

Fig. 1. Comparison between the second difference without average ($z = 0$) and the generalized second difference with $z = 5$, of the spectrum above. The numbers labelling the peaks indicate those peaks identified by the method.

In this representation, S_i may be considered a weighted (unnormalized) average of the data N_j, with C_{ij} representing the weighting factors. The C_{ij} are given by

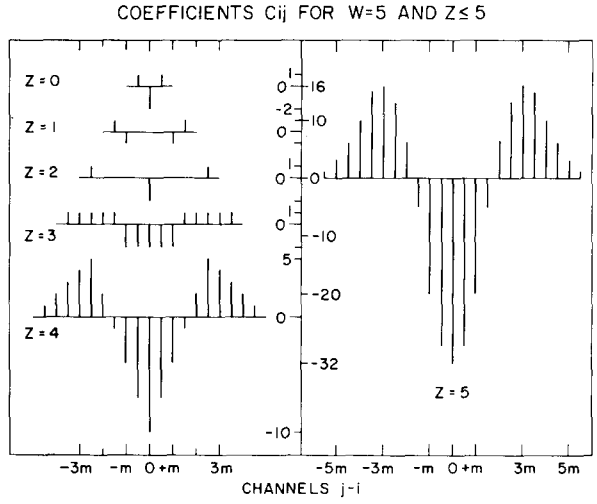

Fig. 2. Weight of each experimental datum N_j as entering in the expression of the generalized second difference $S_i(z, w)$.

$$C_{ij}(z, w) = \sum_{l=i-m}^{i+m} C_{lj}(z-1, w), \qquad (z \geqslant 1)$$

and

$$C_{ij}(0, w) = \begin{cases} 0 \text{ if } |j-i| > 2, \\ 1 \text{ if } |j-i| = 1, \\ -2 \text{ if } \quad j = i. \end{cases} \tag{9}$$

It can be seen that $C_{ij} = 0$ wherever $|j-i| > z \cdot m + 1$.

Fig. 2 shows the weights $C_{ij}(z, w)$ as function of $(j-i)$ for $z \leqslant 5$ and $w = 5$. Although the distributions change with z, there are always two regions of positive weights separated by a region of negative ones such that symmetry with respect to the center is maintained and that $\sum_j C_{ij}(z, w) = 0$. [The latter is easily deduced from the recursion relation (9) since $\sum_j C_{ij}(0, w) = 0$].

This assures that the application of C_{ij} as an operator to a function N_j yields another function S_i which, (a) vanishes if N_j is linear and (b) is similar to the second derivative of a Gaussian function if N_j is a Gaussian plus linear terms, resulting therefore in the preservation of some characteristic geometrical relations. Thus it is possible to look specifically for those portions of the spectrum which are represented by a Gaussian. In particular, this property makes it possible to discriminate against Compton edges.

The effect of the iterative procedure can be interpreted as a modification of the original second difference operator $C_{ij}(z = 0, w)$ such that, if $w \approx \Gamma$, the value of S_{i_0} is essentially given by the difference between N_i's in the vicinity of i_0 and the background intensity

(fig. 2, $z = 5$), rather than the difference between $N_i(i = i_0)$ and its two immediate neighbours, (fig. 2, $z = 0$). Therefore if z is increased up to certain value, S_{i_0} becomes comparatively larger so that weaker peaks can be resolved.

The standard deviation of $S_i(z, w)$ is easily derived from eq. (8) since each term is independent of the others:

$$F_i(z, w) = \left\{ \sum_{\text{all } j} C_{ij}^2(z, w) N_j \right\}^{\frac{1}{2}}. \qquad (10)$$

For practical application in the computer program, the approximation

$$F_i(z, w) = \{\varphi(z, w) N_i\}^{\frac{1}{2}}, \qquad (11)$$

with

$$\varphi(z, w) = \sum_{\text{all } j} C_{ij}^2(z, w)$$

is used. This simpler expression for $F_i(z, w)$ becomes nearly equal to eq. (10) for weak peaks situated on large backgrounds. For intense peaks it is larger than eq. (10) but in these cases the discrepancy is negligible compared with S_{i_0}. Table 1 lists φ values for different z and w values.

In the following we will sometimes omit the parentheses in $S_i(z, w)$ and $F_i(z, w)$ and use simply S_i and F_i.

3. Choice of the parameters z and w

We have obtained a function $S_i(z, w)$ that may be successfully used for peak identification in the presence of background and statistical errors if the parameters z and w are conveniently chosen. To provide a criterion for the quality of $S_i(z, w)$ with respect to the resolution of weaker peaks, the quantity $A_{\min}(z, w)$ has been introduced: the lowest possible value of $A_{\min}$ determines the optimum pair (z_0, w_0).

TABLE 1

Values for $\phi(z, w) = \sum_j C_{ij}^2(z, w)$.

w \ z	1	2	3	4	5
1	6	6	6	6	6
3	4	6	18	82	448
5	4	6	30	350	5220
7	4	6	42	938	27342
9	4	6	54	1974	95034
11	4	6	66	3586	257796
13	4	6	78	5902	592488
15	4	6	90	9050	1209410
17	4	6	102	13158	2258382
19	4	6	114	18354	3934824

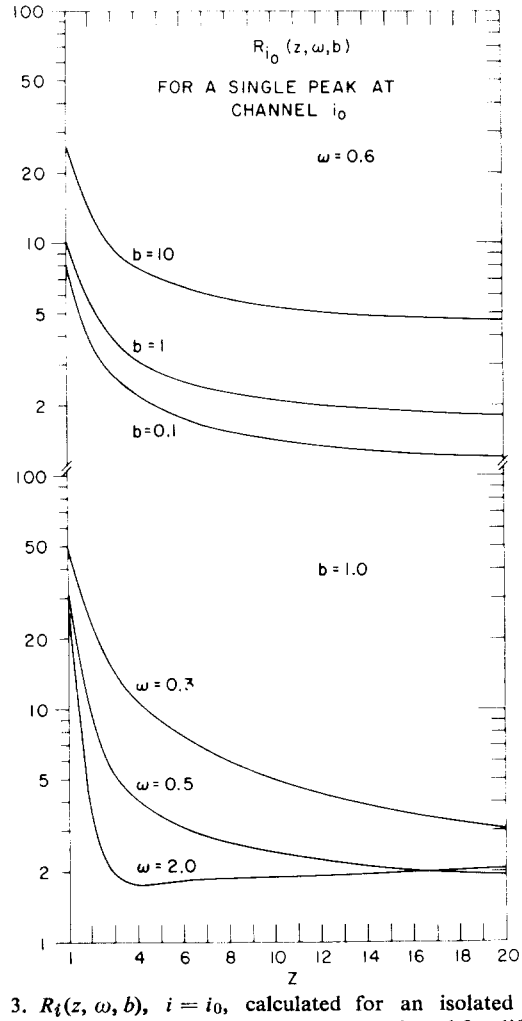

Fig. 3. $R_i(z, \omega, b)$, $i = i_0$, calculated for an isolated peak centered at channel i_0. The upper curves are plotted for different relative background intensities b. In the lower curves the relative window width ω varies while b is constant.

The dependence of $A_{\min}$ upon the parameters z and w is given, implicitly, by the function

$$F_{i_0}/S_{i_0} = f(z, w, \Gamma, A, B, C),$$

which is by definition equal to unity when $A = A_{\min}$. However, in order to reduce the number of variables involved, we will assume in the following analysis $C = 0$ and introduce $b = B/A$ and $\omega = w/\Gamma$. Using these modifications it can be shown that the function

$$R_i(z, \omega, b) = (F_i/|S_i|)(A \cdot \Gamma)^{\frac{1}{2}}, \qquad (12)$$

does not explicitly depend on A and is very nearly independent of Γ for $z \geqslant 2$. It is, hence, preferable to F_i/S_i. From eq. (12) and the definition of $A_{\min}$ we finally obtain

$$A_{\min} = R_{i_0}^2/\Gamma, \qquad (13)$$

whose minimum will coincide with the minimum of R_{i_0}. In fig. 3, $R_{i_0}(z, \omega, b)$ is plotted as a function of z, while ω and b are taken as parameters. In the upper part of the figure $\omega = 0.6$ is fixed and R_{i_0} is shown for different b's. For the lower part of the figure, $b = 1$ and ω ranges from 0.3 to 2.0. The minimum is found to shift toward smaller values of z when ω increases but stops at $z = 4$ if ω is larger than a certain minimum value $\omega_{\min}$ which depends on b. For b ranging from zero to infinity $\omega_{\min}$ varies from 5 to 1. The curves for $\omega \geqslant 2.0$ are almost identical except that increasing values of ω yield a larger positive slope for $z \geqslant 5$ and a more pronounced minimum at $z = 4$. Since the minimum value of R_{i_0} increases only slightly with ω, we conclude that for the case of a single peak, $z = 4$ and $\omega \geqslant 5$ is the best combination for any intensity of the background.

However, the result obtained from the analysis of an isolated peak is not sufficiently general since also doublets can appear in the experimental spectra.

Unlike a single peak, a doublet may not be treated with a large window width without destroying the individuality of its components. This can be shown with the aid of fig. 4 where a particular doublet has been drawn as well as its associate S_i function. Let us consider the value of $S_i(z, w)$ which must be positive at $i = i_m$ to make the recognition of a doublet possible. Assuming an "average" of this curve is carried out with w larger than the distance between peak centers,

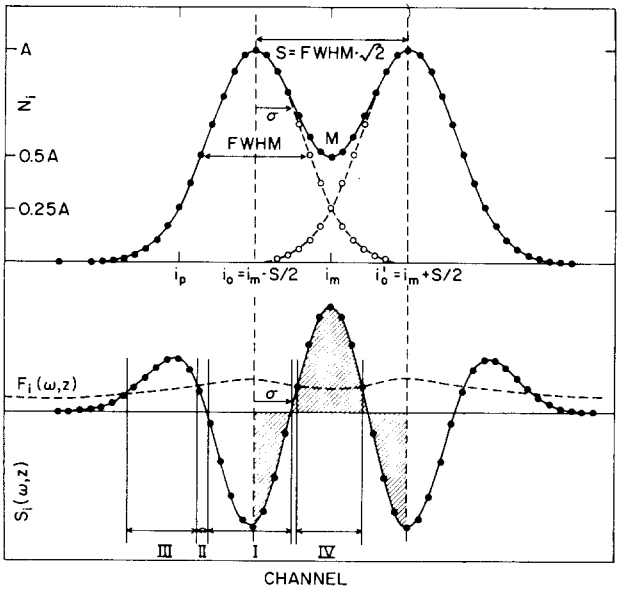

Fig. 4. A doublet with separation between peak centers s = fwhm $\sqrt{2}$ and its associate $S_i(z, w)$ function. The number of channels in region I, II and III (or IV) are taken into account to identify a set of points as a peak.

the new value of $S_i(z+1, w)$ at $i = i_m$ will be negative since the sum of the hatched areas is zero. Therefore if the distance between centers is of the order of Γ, it is necessary that $\omega \leqslant 1$. This can be shown quantitatively, as follows:

We consider a new "number of counts" function N_i, to describe a doublet which consists of the sum of two Gaussians plus the constant background assumed above. In order to keep the same parameters used in the analysis of a single peak we will take a particular doublet (fig. 4) such that; a) both peaks have the same amplitude; b) the separation between peaks is fixed so that the central point M be at half maximum. Then

$$N_i = A\,[\exp(i-i_m+\tfrac{1}{2}s)^2+\exp(i-i_m-\tfrac{1}{2}s)^2+b]\,,$$

where $s = \Gamma\cdot\sqrt{2}$.

In this case R_i, eq. (12), was calculated at $i = i_m$. Its minimum defines a new pair (z, w) which in turn through eq. (13) yields $A_{\min}$ (doublet).

The function R_{i_m} tends to infinity when z approaches the asymptotic value $z_a(\omega)$ for which $S_{i_m} = 0$. Values of $z > z_a(\omega)$ yield $S_{i_m} < 0$ and therefore are not applicable since the method demands at least one "positive" point (and larger than its standard deviation) as the initial indication for the presence of a peak. (section 4).

The results for $z < z_a(\omega)$ are shown in fig. 5 in a similar way as in fig. 3 to ease the comparison. It is seen that possible z values are more restricted as ω increases and further, no positive $S_i(i = i_m)$ is obtained if $\omega \gtrsim 1$. This result is contrary to the one obtained in the case of a single peak.

As was shown above, the choice of the most suitable pair (z, ω) permits identification of the weakest peaks the method can recognize, because in this case $A_{\min}$ takes on its smallest possible value. However the setting that gives the best result for an isolated peak may be such that a weak doublet may be ignored, and vice versa. In order to establish a choice of z and ω suitable for a real spectrum and in view of the opposite results regarding ω for singlets and doublets we used the following criterion. We choose as optimum combination (z_0, ω_0) the one that minimizes the product of $A_{\min}$ (singlet) and $A_{\min}$ (doublet) so that reasonable results can be simultaneously expected for both cases.

Because $A_{\min}$ is related to R_i by eq. (13), the pair (z_0, ω_0) also makes the product R_{i_0} (singlet) $\times$ R_{i_m} (doublet) a minimum. This product is plotted in fig. 6. For any b, the smallest absolute value is found at $z_0 = 5$ and $\omega_0 = 0.6$ ($w_0 = 0.6\,\Gamma$), as reported in the previous section. Accordingly the example of fig. 1 was

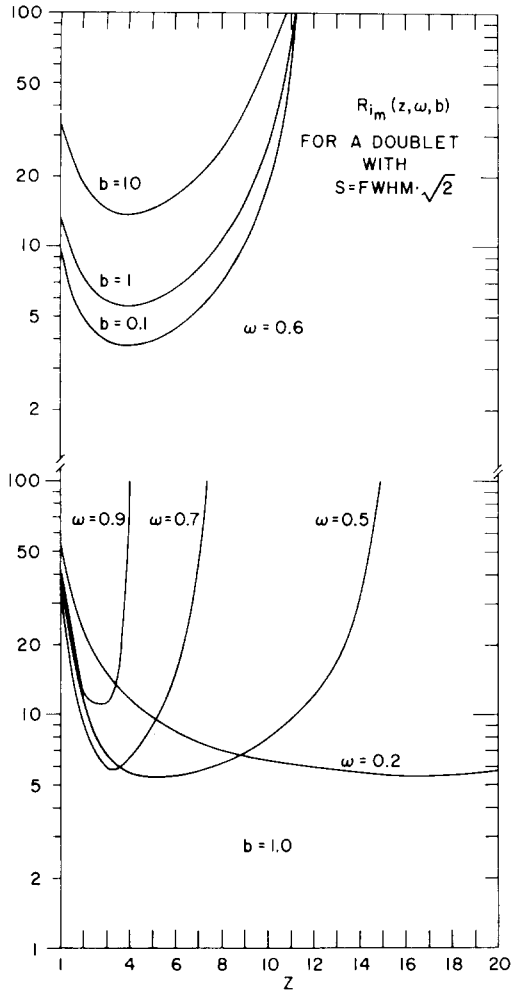

Fig. 5. $R_i(z, \omega, b), i = i_m$, calculated for a doublet centered at channel i_m. The upper curves are plotted for different relative background intensities b. In the lower curves the relative window width ω varies while b is constant.

treated with $w = 5$ since Γ covers about 8 channels. It may be noted that even the component peaks in the marginal doublet near channel 600 are well resolved.

Another example of the analysis of a spectrum at lower energy and $\Gamma \simeq 10$, ref.[3]), is presented in fig. 7, with $w = 7$.

4. Application of $S_i(z_0, w_0)$ to a peak finding procedure

If the background were strictly a succession of straight sections throughout the spectrum, the presence of a peak could be determined by the condition $S_i > fF_i$ where f is just a factor of confidence. Sometimes, however, as in the case of Compton edges, the background is far from linear. It is then necessary to impose additional conditions, based on some distinctive peak characteristics, which must be met before a peak is recognized. In the following the adopted procedure to achieve this is explained.

The computer uses the function $S_i(z_0, w_0)$ to look for sets of channel numbers $i_1 \leqslant i_2 \leqslant i_3 \leqslant i_4 < i_5$ as defined in table 2.

When such a set is found we assume that it represents a peak if S_i has certain features which can be expressed in terms of relations among $i_1, i_2, \ldots, i_5$ and which are characteristic of the second derivative of a Gaussian function. These relations are contained in the four following conditions. The computer recognizes a peak whenever these conditions are fulfilled.

The first condition is

$$|S_{i_4}| > 2F_{i_4}. \tag{14}$$

The second condition concerns the expected number of negative points n_1 in region I of fig. 4, depending

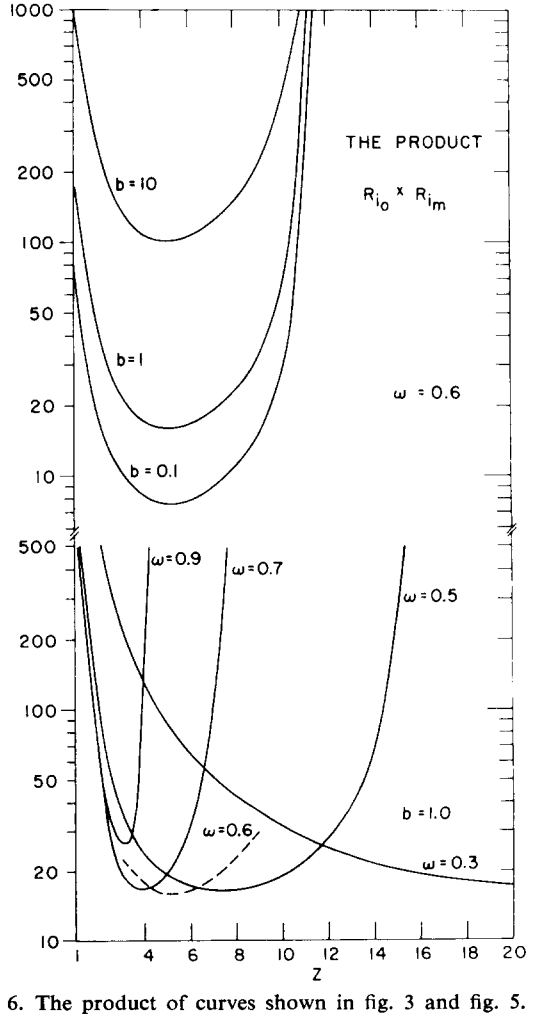

Fig. 6. The product of curves shown in fig. 3 and fig. 5. The absolute minimum of this product is found at $z_0 = 5$ and $\omega_0 = 0.6$, for any b.

on the value of Γ, w and z. To evaluate it we again assume that the data are a continuous function of channel x, eq. (1). Then

$$G_0''(x) = \{(x^2-\sigma_0^2)/\sigma_0^2\}\, G_0(x), \qquad (15)$$

will be used in place of $S_i(z = 0)$. Here $G_0(x)$ represents a Gaussian as in eq. (1), with $\bar{x} = 0$ and $\sigma = \sigma_0$.

To obtain the analogous function of $S_i(z = 1)$, we integrate $G_0''(x)$ within an interval w

$$G_1''(x) = \int_{x-\frac{1}{2}w}^{x+\frac{1}{2}w} G_0''(x)\, dx. \qquad (16)$$

Introducing a new width σ_1 we assume that in the vicinity of $x = \sigma_1$, $G_1''(x)$ can be expressed in the form of eq. (15). Hence σ_1 is defined by the condition that $G_1''(x)$ vanishes for $x = \sigma_1$, or equally by

$$G_0'(\sigma_1+\tfrac{1}{2}w) = G_0'(\sigma_1-\tfrac{1}{2}w), \qquad (17)$$

where $G_0'(x)$ stands for the first derivative of $G_0(x)$. Eq. (17) assumes the form

$$\log\{(1+\tfrac{1}{2}w/\sigma_1)/(1-\tfrac{1}{2}w/\sigma_1)\} = w(\sigma_1/\sigma_0^2),$$

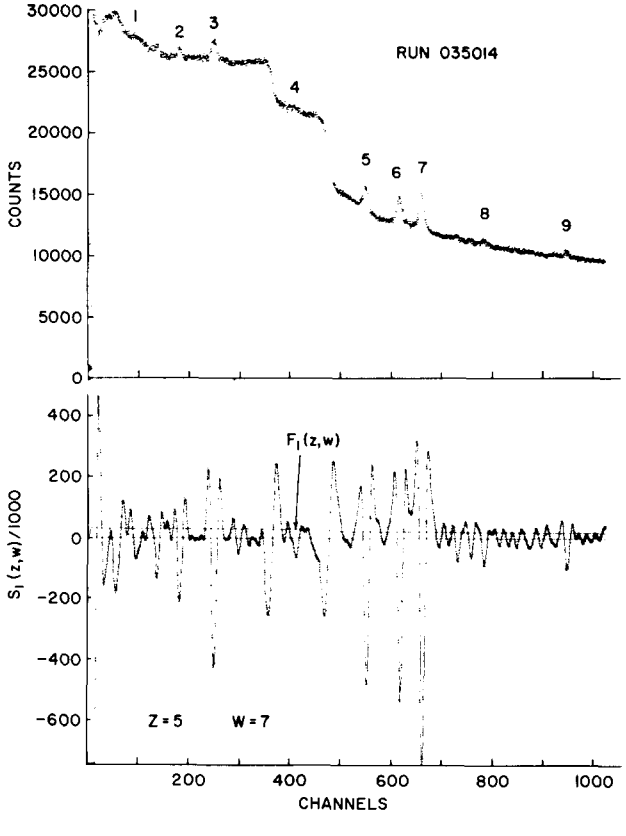

Fig. 7. Another example similar to fig. 1, of the applicability of the method, showing the discrimination between peaks and Compton edges.

TABLE 2

Definition of the channel numbers used to recognize a peak.

Channel	Definition	Description (fig. 4)
i_1	$S_{i_1} > F_{i_1}$ $S_{i-1} < F_{i-1}$	First channel in region III and in region IV.
i_2	$S_{i_2+1} < F_{i_2+1}$ $S_j > F_j \quad (i_1 \leqslant j \leqslant i_2)$	Last channel in region III and in region IV.
i_3	$S_{i_3} < 0$ $S_j > 0 \ (i_1 \leqslant j \leqslant i_3-1)$	First channel in region I.
i_4	$S_{i_4} \leqslant S_j \quad (i_3 \leqslant j \leqslant i_5)$	Coincident with i_0 in the particular case shown in fig. 4.
i_5	$S_{i_5+1} > 0$ $S_j < 0 \quad (i_3 \leqslant j \leqslant i_5)$	Last channel in region I.

which, using the expansion of the logarithmic function up to the fourth term, reduces to

$$(-\sigma_1^4/\sigma_0^2)+\sigma_1^2+\tfrac{1}{12}w^2 = 0.$$

The only real positive root of this equation is

$$\sigma_1 = f_1\,\sigma_0,$$

with

$$f_1 = [\tfrac{1}{2}+\tfrac{1}{2}(1+\tfrac{1}{3}w^2/\sigma_0^2)^{\frac{1}{2}}]^{\frac{1}{2}}.$$

Once $G_1''(x)$ is known, eq. (16) can be applied in the same way to get

$$\sigma_2 = f_2\,\sigma_1 = f_2 f_1\,\sigma_0.$$

In general, we have

$$\sigma_z = f_z\,\sigma_{z-1} = f_z \ldots f_2 \cdot f_1\,\sigma_0 = f_z'\,\sigma_0$$

and the expected number of negative points in region I is taken as the integer number nearest to the value

$$2\,\sigma_z = 2 f_z'\,\sigma_0 = 0.849 f_z' \cdot \Gamma = k_z \cdot \Gamma,$$

or

$$n_1 = |[k_z \cdot \Gamma + \tfrac{1}{2}]|, \qquad (18)$$

where the symbol $|[a]|$ denotes integer part of a.

Fig. 8 is a plot of $k_z = k_z(\omega)$ for three different z's.

The value of n_1 enables us to arrive at the second condition

$$i_5 - i_3 + 1 = n_1 \pm \varepsilon. \qquad (19)$$

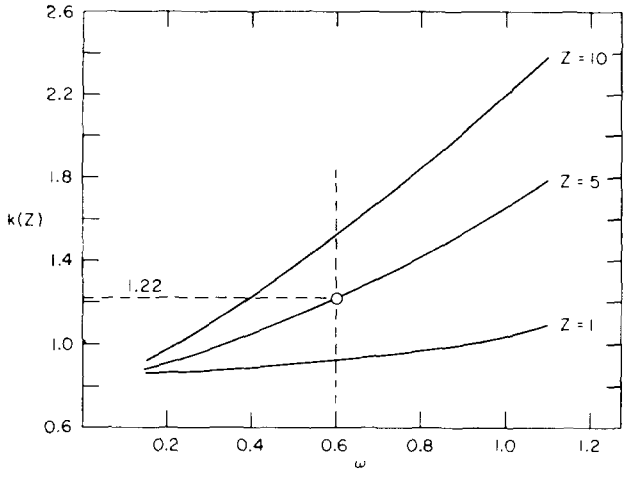

Fig. 8. The proportional factor $k(z)$ between fwhm, the expected number of channels n_1 in region I (fig. 4).

Here ε stands for the tolerance admitted in the agreement with n_1. The deviations from this expected number are mainly due to the uncertainty in Γ which in addition may vary along the spectrum. For this reason it is necessary to allow for differences in n_1.

The analysis of many spectra has shown that a tolerance of $\varepsilon = 2$ is reasonable and still retains the usefulness of condition (19). It can be added that a mistaken identification of a peak is further checked by the fitting subroutine itself. If a reasonable fitting is not accomplished, the line is disregarded. In this sense it is always convenient to choose a wide limit rather than to take the trouble of carefully determining for each case the possible variations of Γ.

The third and fourth conditions concern the number of channels in regions II and III, fig. 4. The use of eq. (15) could be extended to calculate these numbers but this would lead to a new transcendental equation whose solution requires a somewhat complicated computation. The application of these quantities, however, does not need such accuracy and the following approximations, based on n_1, are sufficient. The expected number of channels in region II

$$n_2 = |[(F_{i_0}/S_{i_0})\,\tfrac{1}{2}\,(n_1 \pm \varepsilon)+\tfrac{1}{2}]|\,, \qquad (20)$$

is used for the third condition

$$i_3 - i_2 - 1 \leqslant \begin{cases} n_2 & \text{if } n_2 \geqslant 1, \\ 1 & \text{if } n_2 = 0. \end{cases} \qquad (21)$$

Similarly, to obtain the last condition it is assumed that the positive region of $G''(x)$, eq. (15), is a triangle of base $n_1 \pm \varepsilon$ and height $\frac{1}{2}\,S_i(i = i_0)$. Then

$$n_3 = |[(n_1 \pm \varepsilon)\cdot\{1-2\,(F_{i_0}/S_{i_0})\}+\tfrac{1}{2}]|, \qquad (22)$$

is the expected number of channels in region III. While n_1 and n_2 are valid in any case, n_3 does not apply for the recognition of the second peak of a doublet as it can be seen by considering region IV in fig. 4. The expected number of channels n_4 in region IV is in general different from n_3 because it depends on the separation between the two peaks. On the other hand, however, both regions III and IV, are considered equivalents by the method according to table 2. Assuming that the minimum separation that permits the resolution of a doublet is $s = \Gamma\sqrt{2}$, for which $n_4 \approx n_3$, the fourth condition will demand, then, that the number of channels in either region III or IV, must be always larger than n_3. Hence the fourth condition is

$$i_2 - i_1 + 1 \geqslant n_3. \qquad (23)$$

In summary, the computer calculates S_i, F_i, n_1, n_2, and n_3. The only datum needed is Γ. Using these quantities it searches for sets $\{i_1 \cdots i_5\}$ that fulfill the above conditions. These sets represent peaks and for each one a least squares fit is carried out as explained in the next section.

In the spectra of fig. 1 and fig. 7, the peaks identified by this method are numbered, to give an example of its applicability. In particular they show the results obtained with doublets (fig. 1) and the ability to discriminate between peaks and Compton edges (fig. 7).

5. Peak fitting and results

A general non-linear least squares fitting subroutine[2]) is added to the program to determine the centroid, amplitude and width of each peak. The subroutine fits a function $f(i;\, p_j)$ of certain parameters p_j to a series of experimental points N_i. It uses the Newton-Raphson method of iteration to find the j-component vector $\boldsymbol{p}$ which minimizes the quadratic form

$$Q(\boldsymbol{p}) = \sum_i [f(i;\, \boldsymbol{p}) - N_i]^2/N_i \qquad (24)$$

by linearization and iteration. At every step k, one obtains a vector $\boldsymbol{p}_k$ and a value $Q_k = Q(\boldsymbol{p}_k)$.

The condition which terminates iteration is either

$$|Q_k - Q_{k-1}| < \delta\, Q_0, \qquad (25)$$

where δQ_0 is a measure of the demanded accuracy or

$$k > k_{\max}. \qquad (26)$$

If the second condition terminates iteration, this usually

means that a fitting has not been achieved (the iteration procedure did not converge).

The subroutine must therefore be supplied with δQ_0, k_{max}, an initial estimate of p_j, the function $f(i; p_j)$ and the data N_i, ($i_{min} < i < i_{max}$), to be fitted.

The first two limit values are taken as $\delta Q_0 = 0.1$ and $k_{max} = 15$.

For the initial estimate of the peak position i_0, the principal parameter, we take the center of gravity of region I, fig. 4. This is

$$i_0 = \sum_{S_i<0} i \cdot S_i / \sum_{S_i<0} S_i. \tag{27}$$

With both i_0 and an estimated width $(\Gamma)_{est}$ (see below), we define the channels that will undergo the fitting:

$$i_{min} = i_0 - 5(\Gamma)_{est},$$

$$i_{max} = i_0 + 5(\Gamma)_{est}.$$

If the following peak is at channel i_0' and if condition

$$i_0' < i_{max}, \tag{28}$$

is fulfilled, then both lines will be considered as a doublet. In either case, we use the following function to fit the data:

$$f(i; p_j) = p_1 \exp(-t_{23}^2) + p_4 \exp(-t_{53}^2) + \\ + p_6 + p_7 x + p_8 x^2, \tag{29}$$

where $p_4 \equiv 0$ if the peak is single, $p_4 \neq 0$ if there is a doublet and

$$t_{mn} = 1.665\,(i - p_m)/p_n.$$

The initial estimate of the parameters which feed the subroutine are taken as

$$p_1 = N_{i_4} - P_6,$$

$$p_2 = i_0,$$

$$p_3 = (\Gamma)_{est},$$

$$p_4 = \begin{cases} 0, & \text{if a single peak,} \\ N_{i_4}' - p_6, & \text{if a doublet,} \end{cases}$$

$$p_5 = i_0', \tag{30}$$

$$p_6 = \left(\sum_{i=i_{min}-l}^{i_{min}+l} N_i + \sum_{i=i_{max}-l}^{i_{max}+l} N_i \right) / (4l+2),$$

$$p_7 = \left(\sum_{i=i_{max}-l}^{i_{max}+l} N_i - \sum_{i=i_{min}-l}^{i_{min}+l} N_i \right) / \{(2l+1)(i_{max} - i_{min} + 1)\}$$

where $l = 1$ is usually appropriate,

$$p_8 = 0.$$

When the iteration has converged, these parameters represent the result of the analysis for each peak.

So far, the only quantity that must be provided is Γ, which is used in turn to define w. The peak finding procedure needs Γ to calculate S_i and F_i, using eq. (7) and n_1, eq. (18). The value of Γ is also needed to establish both the channels that will enter the fitting subroutine and the initial estimate p_3, eq. (30), for each peak. In practice a rough estimate of Γ would be enough except for p_3. An improved estimate as it is required by the fitting subroutine makes necessary a previous plot of the data. In addition, the value of Γ depends on the energy and sometimes more than one estimated value for Γ should be provided. To avoid the undesirable delay that this would imply, we have introduced the following simplification. From a crude estimate of Γ, the program is supplied with w according to eq. (7). Thus S_i, F_i and $n_1 = k_z\, w/0.6$ are obtained. On the other hand the fitting subroutine uses a value of Γ which is self-determined by the program through an iterative procedure. A fit is first attempted with $\Gamma = 2$. If the iteration does not converge, Γ is increased by a fixed amount and so forth, until a maximum $(\Gamma)_{max}$ is reached. This procedure has been successfully applied. The required increase in computer time is negligible and no precise knowledge of Γ is necessary to run the program.

The computer supplies a listing of the results, as shown in table 3, for the spectra of fig. 1 and fig. 7.

A row is filled with zeros when both $k > k_{max}$ and $\Gamma > (\Gamma)_{max}$ before the fitting is accomplished (as for peak 8 in the first listing and peak 1 in the second listing of table 3).

The ninth column gives χ^2, eq. (24), for the last iteration; the tenth column gives the number of that iteration. The last column shows the number $i_{max} - i_{min} + 1$.

In general $p_4 = 0$ in eq. (29), i.e., a singlet is fitted. Because in eq. (29), the width Γ (represented by p_3), is the same in both exponentials, two identical adjacent Γ values and associate errors in columns 7 and 8 of table 3, indicate those cases in which $p_4 \neq 0$. Examples are the pairs of peak number 9–10 and 12–13 in the upper listing of the same table.

As an optional output the program also plots for

TABLE 3

Peak no.	Position and error		Height and error		Relative	fwhm and error		χ^2	No. of last iter.	No. of points
	Position, intensity and width of the peaks in the run 145014 (fig. 1)									
1	79.059	0.096	3928.	128	1.0000	7.18	0.29	73.	6	41
2	183.941	0.582	312.	77	0.0794	5.23	1.62	29.	11	41
3	208.757	0.926	177.	124.	0.0451	2.97	2.63	18.	6	21
4	229.918	0.335	862.	94.	0.2195	9.58	1.25	38.	5	41
5	313.360	0.851	239.	56.	0.0608	8.20	2.41	41.	9	61
6	379.534	0.348	749.	84.	0.1908	9.14	1.25	36.	7	41
7	473.663	0.520	276.	130.	0.0703	2.42	1.42	19.	6	21
8	0.	0.	0.	0.	0.	0.	0.	0.	0	0
9	591.279	0.620	401.	78.	0.1020	6.96	1.72	74.	9	60
10	609.569	0.667	371.	81.	0.0945	6.96	1.72	74.	9	60
11	847.729	0.544	236.	84.	0.0601	3.22	1.40	37.	7	41
12	893.735	0.179	1437.	69.	0.3659	9.50	0.58	64.	7	59
13	912.424	0.360	696.	81.	0.1773	9.50	0.58	64.	7	59
	Position, intensity and width of the peaks in the run 035014 (fig. 7)									
1	0.	0.	0.	0.	0.	0.	0.	0.	0	0
2	182.259	0.431	775.	107.	0.3022	8.63	1.49	33.	5	41
3	249.857	0.290	1610.	168.	0.6269	10.96	1.24	47.	15	41
4	409.863	0.411	446.	212.	0.1739	1.63	0.87	24.	4	21
5	549.883	0.200	1678.	112.	0.6546	10.16	0.78	49.	7	41
6	616.378	0.215	1793.	124.	0.6992	9.70	0.80	73.	7	41
7	660.030	0.097	2564.	81.	1.0000	10.12	0.37	29.	8	41
8	782.065	0.494	311.	86.	0.1214	3.84	1.30	40.	3	41
9	946.043	0.797	327.	84.	0.1276	8.46	2.71	55.	8	41

each peak the calculated function $f(i, \boldsymbol{p}_j)$ (p_j here are the final parameters) superimposed on the experimental data that it fits. Fig. 9 shows the plot obtained for some of the peaks in the spectrum of fig. 1. The computer makes these plots automatically, labelling them and choosing adequate scales.

6. Summary

A flow chart of the entire program is shown in fig. 10. It summarizes the automatic analysis procedure mentioned in the introduction.

The program reads the experimental data stored on paper or magnetic tape as they come from the analyzer. Then it reads w. With w and $z = 5$, the functions S_i and F_i are computed as in eq. (6) and eq. (11), respectively. Next it calculates n_1, n_2 and n_5 using $\Gamma = w/0.6$ in eq. (18) and searches for those sets $\{i_1, i_2, \ldots i_5\}$, (defined in table 2) which fulfill the conditions (14), (19), (21) and (23).

When the whole spectrum has been scanned and the parameter i_0 is determined and stored for each peak found, the program takes one peak at a time to initiate the fitting procedure.

Starting with the minimum $\Gamma = 2$, the channels to undergo fitting are established and the initial parameters p_j are calculated, eq. (30), according to condition (28).

The subroutine stops when either condition (25) or (26) is reached. In the first case the results are stored and the next peak is taken to restart the cycle. The second possibility leads back to the subroutine with a new increased value for Γ. If Γ exceeds the established maximum without a fit having been achieved, the peak involved is disregarded and the cycle restarts with the following one. When it fails to fit a doublet, the second component is considered as the next peak. Finally the two types of output supplied, a plot and a listing, are indicated in fig. 10.

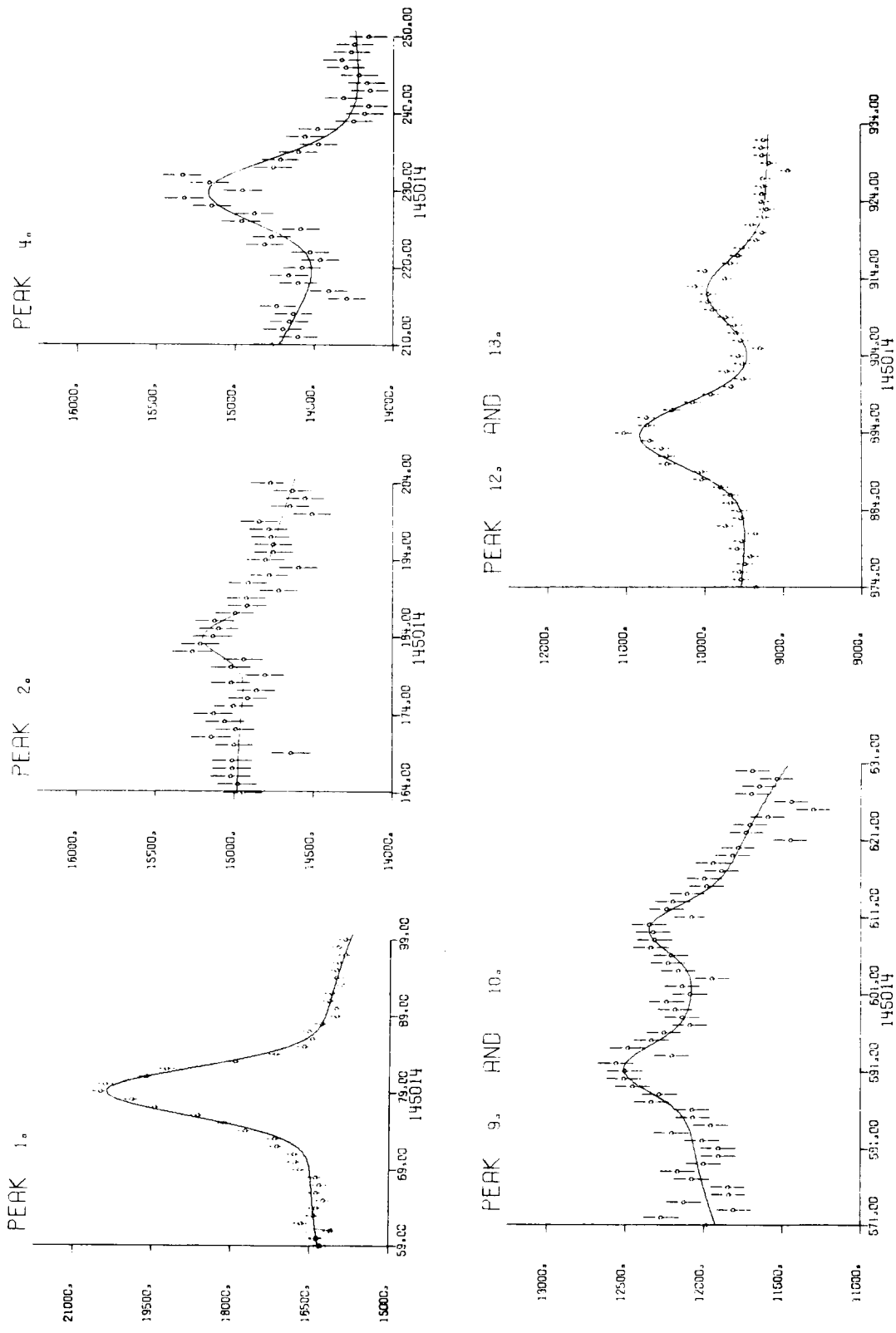

Fig. 9. Plot of some of the peaks in the spectrum of fig. 1 and the function which fits each one best. The computer makes these plots automatically, labeling them and choosing the adequate scales.

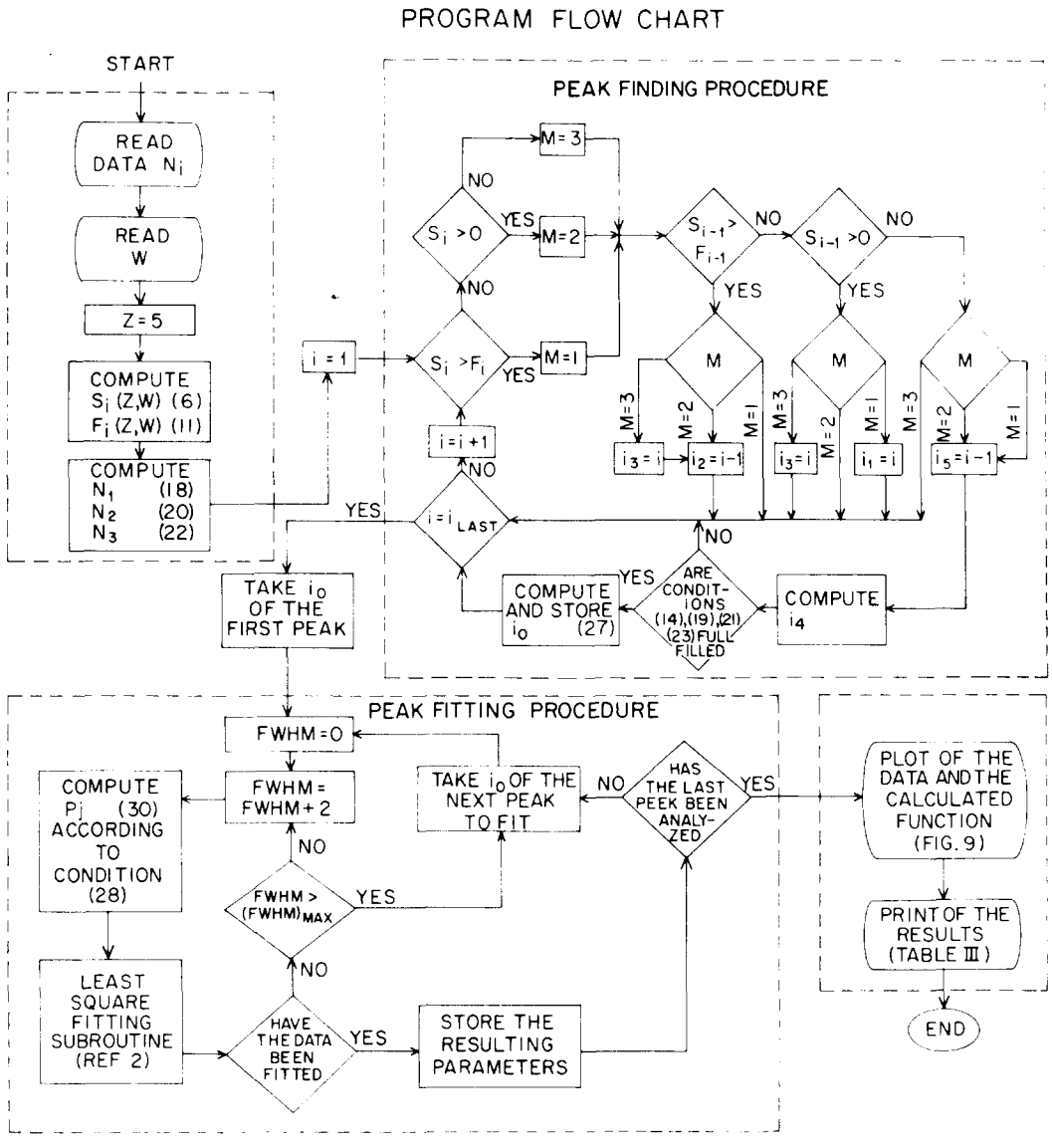

Fig. 10. Flow chart of the entire program. The different blocks framed in dashed lines are: 1. Input and previous computations; 2. Peak finding procedure; 3. Peak fitting procedure; 4. Output.

The author wishes to thank Dr. G.S. Goldhaber who kindly provided him the very valuable opportunity of working in the Nuclear Structure Group at this Laboratory. The fellowship granted to him by the Consejo Nacional de Investigaciones Cientificas y Tecnicas, Argentina, during the past year is also greatly appreciated. He is particularly indebted to Dr. Salvador Bozzo whose ideas and enthusiasm have largely contributed to originate this work, and to Drs. H. Paul and P. Thieberger for valuable suggestions and assistance. The stimulating support received from Drs. C. Chasman, G. Emery, W. Kane and R. Ristinen is gratefully acknowledged.

References

1) Application of Computers to nuclear and radiochemistry, National Academy of Sciences, National Research Council — NAS-NS 3107 (1962).

2) E.A. Crosbie and J.E. Monahan, A general least squares fitting procedure, AN E208, Argonne National Laboratory (1959).

3) G.T. Emery, W.R. Kane and M. Mariscotti, Measurements of gamma rays from the $^{53}Cr(n, \gamma)^{54}Cr$ reaction, to be published.